P2 diversity data guide

Created by: Michael Van Nuland (March 2017)

This is a guide to import, analyze, and generate figures with Project 2 data on *Solidago* spp. responses to water (low = 80 mL, high = 160 mL) and diversity (monoculture vs. mixture) treatments. Although the steps in this guide are less detailed than the P1 tutorial, they provide enough raw material (code) for you to test your hypotheses. Please note, this guide uses an example data set, which you can swap in for the real one once it has been posted on blackboard. **Please contact me (Michael) sooner than later if you're stuck or have questions.**

STEP 1. Download the "P2_div_data_EXAMPLE.csv" file from blackboard and **save it to your desktop**. (Do the same for the real file once it is uploaded on blackboard).

STEP 2. Open RStudio and set the working directory to your desktop. Do this by going to Sessions (top menu bar) -> Set working directory -> Choose directory -> Desktop

STEP 3. Read in the data file and assign it to a "P2_div_data" object. Because you've set your working directory to your desktop, and that's where you downloaded the data to, there should be no issues. If you get an error, go back and make sure you did STEP 1 and STEP 2 succesfully.

```
P2_div_data <- read.csv(file="P2_div_data_EXAMPLE.csv")    # The real data will be "P2_div_data.csv"
```

STEP 4. Take a look at the first lines of the data to get oriented. Notice how the species are grouped whether they are generally adapted to dry (xeric) or wet (mesic) conditions. We then have the H2O treatments, diversity treatments, initial trait data (t0), final trait data (t35), and above- and belowground biomass data for the focal plants (e.g., "Focal_Aboveground_biomass"). We also have a column called "Ecosystem_function", which is the combined total biomass of all 6 plants in each pot (biomass of focal plant + biomass of the 5 other plant species in the pot).

```
head(P2_div_data)
```

STEP 5. Calculate relative growth rate (RGR) for height, root-to-shoot ratio, and add both to the P2_div_data object. NOTE: you can do the same for the leaf number data, just make a different object than "RGR" (maybe something like "Leaf_RGR")

```
RGR <- ((log(P2_div_data$Height_t35) - log(P2_div_data$Height_t0))/(35 - 0))

Root_shoot <- (P2_div_data$Focal_Aboveground_biomass/P2_div_data$Focal_Belowground_biomass)

P2_div_data <- cbind(P2_div_data, RGR, Root_shoot)

head(P2_div_data)  # Double check that there are RGR and Root_shoot column at the end of the file
```

```
##      Species H2O_adapted H2O_treatment Div_treatment LeafNumber_t0
## 1 S. bicolor       Xeric           Dry           Mix             4
## 2 S. bicolor       Xeric           Dry           Mix             4
## 3 S. bicolor       Xeric           Dry           Mix             4
## 4 S. bicolor       Xeric           Dry          Mono             4
## 5 S. bicolor       Xeric           Dry          Mono             4
## 6 S. bicolor       Xeric           Dry          Mono             4
##   Height_t0 LeafNumber_t35 Height_t35 Focal_Aboveground_biomass
## 1       7.2              6        7.3                    433.03
## 2       4.7              6        9.1                    347.50
## 3       4.7              6        8.4                    297.22
## 4       7.5              8        9.4                    423.07
```

```
## 5       7.9              8        9.7                    467.12
## 6       5.5              8        8.5                    368.66
##   Focal_Belowground_biomass Ecosystem_function          RGR Root_shoot
## 1                    207.94           597.0884 0.0003940949   2.082476
## 2                    301.43           588.5176 0.0188774830   1.152838
## 3                    285.12           533.6575 0.0165905485   1.042438
## 4                    329.42           364.2466 0.0064516191   1.284288
## 5                    302.75           415.7178 0.0058646607   1.542923
## 6                    285.49           370.2059 0.0124376592   1.291324
```

STEP 6. Generate your hypotheses and test them with two-way ANOVAs. This is the most appropriate test for these data to look at both main effects and interaction effects. Example code is below (be sure to switch "PLANT_TRAIT" for an actual trait column name):

```
Hyp1 <- lm(PLANT_TRAIT ~ Div_treatment * H2O_treatment, data = P2_div_data)

anova(Hyp1)

TukeyHSD(aov(PLANT_TRAIT ~ Div_treatment * H2O_treatment, data = P2_div_data))
```

Remember that a Tukey's HSD post-hoc test is important to see which groups and treatments are different from one another. Comparisons that are different (left-most "diff" column) will have significant p-vales (right-most column).

You may need to separate the data by different treatment groups (for example, by *Solidago* spp., by watering treatments, etc) to test different hypotheses. **The data is currently arranged to easily separate by species** with the following:

```
XericSpp_P2_div_data <- P2_div_data[1:12, 1:13]  # Take the first 12 rows that is the xeric species (S.

MesicSpp_P2_div_data <- P2_div_data[13:24, 1:13]  # Ditto for the last 12 rows and mesic species (S. ri
```

This code makes new objects that you assign a subset of the P2_div_data object to. The code between the brackets says to take data from a subset of rows, and all the columns. So for the xeric species (*Solidago biocolor*), the code is saying take rows 1 though 12, and all the columns (1 through 13); for the mesic species, it says take the other rows (13 though 24) and all columns. Note that if your P2_div_data as more or less than 13 columns, make sure you alter that second set of numbers in the brackets to reflect that.

The easiest way to separate out the dataset into the other treatment groups is to:

1) open the data file in excel

2) sort the data by the treatment group your interested in

3) save (as .csv), read-in that new file (**Step 2**), and assign it (**Step 3**) to a different object (e.g., P2_div_data_sorted)

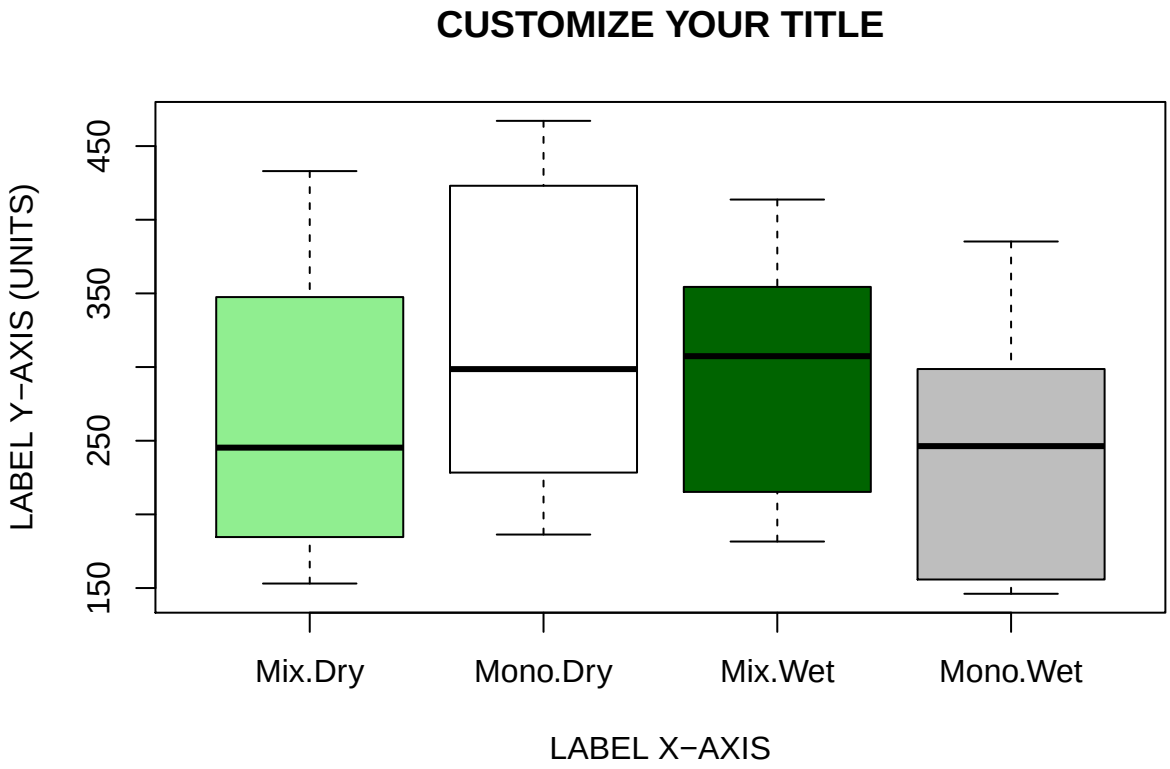
4) look at the object and see which rows correspond to the different treatments. It should be that the first 12 correspond to one group, and the last 12 correspond to the other group.

5) make new objects with the subset of data corresponding to the treatment groups you're interested in separating by.

Step 7. Make figures associated with each of your hypotheses. Before, we had you make boxplots such as with the following:

```
boxplot(Focal_Aboveground_biomass ~ Div_treatment*H2O_treatment, data = P2_div_data,
      col=(c("lightgreen","white", "darkgreen", "grey")),   # greens = mixture, white/grey = monocultur
```

```
      main="CUSTOMIZE YOUR TITLE", xlab="LABEL X-AXIS", ylab="LABEL Y-AXIS (UNITS)")
```

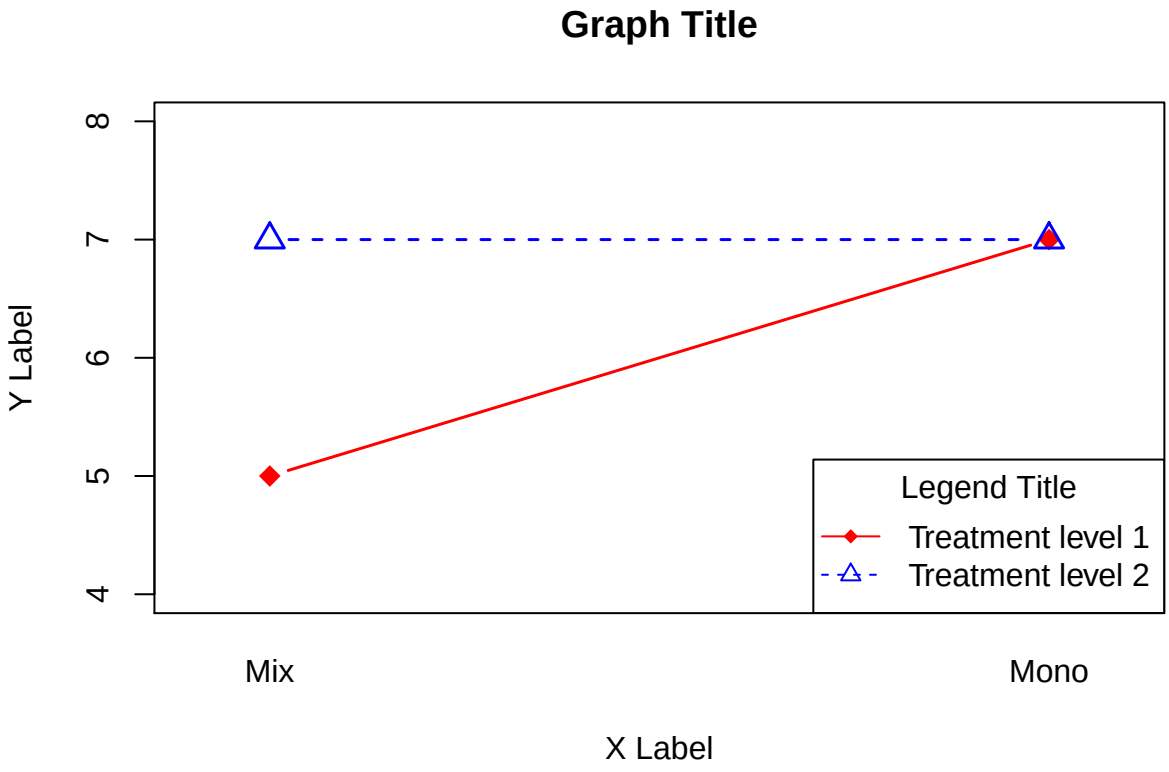


A different (and often better) way to visualize the results is with an interaction plot. The general code format is simply:

interaction.plot(factorA, factorB, Response variable)

An example is plotted below. Note that you need to highlight and run the entire chunk of code below together for it to plot correctly with the appropriate legend.

```
interaction.plot(P2_div_data$Div_treatment, P2_div_data$H2O_treatment, P2_div_data$LeafNumber_t35,
                 type="b", col=c("red","blue"), lty=c(1,2), lwd = 1.5,
                 pch=c(18,24), cex = 1.5, ylim = c(4,8),
                 xlab="X Label", legend=FALSE, ylab="Y Label", main="Graph Title")
legend("bottomright", c("Treatment level 1","Treatment level 2"),
       col=c("red","blue"), lty=c(1,2), pch=c(18,24), title="Legend Title")
```



Any customizations to points, lines, and colors should also be copied to the legend portion beneath the main figure code.

Here's a quick cheatsheet to get you started customizing your plots:

col = color, and should be specified as "col = c(color1, color2)"
lty = linetype, and should be specified as "lty = c(#, #)"
lwd = line width
pch = point characters (different shapes)
cex = point size
ylim = lets you set a specific range of values on the y-axis

STEP 7. Add your specific hypotheses, statistics (F values and p-values), and figures to your papers. Remember, you should explore patterns in all the different traits, but that doesnt mean you should include all the stats or every possible figure. It's best to examine all the results and then "find the story" in the data - what are the clearest patterns? what things are surprising? what things are novel? what things are important in the broader context of increasing drought? plant-microbe interactions? etc.

In your results, be precise and mention the effect size that different treatments had.

Here's a **bad example**: "The mesic species responded differently to water treatments than the xeric species."

Here's a **good example**: "The mesic species (S. ridellii) grew 35% taller than the xeric species (S. bicolor) in the high water treatment, but this relationship was opposite in the low water treatment. Specifically. . . "